Measuring characteristics of individuals: An updated systematic review of instruments' psychometric properties

Cameo Stanick[1] (ID), Heather Halko[2], Kayne Mettert[3] (ID), Caitlin Dorsey[3], Joanna Moullin[4], Bryan Weiner[5], Byron Powell[6] (ID) and Cara C Lewis[3] (ID)

Abstract

Background: Identification of psychometrically strong implementation measures could (1) advance researchers' understanding of how individual characteristics impact implementation processes and outcomes, and (2) promote the success of real-world implementation efforts. The current study advances the work that our team published in 2015 by providing an updated and enhanced systematic review that identifies and evaluates the psychometric properties of implementation measures that assess individual characteristics.

Methods: A full description of our systematic review methodology, which included three phases, is described in a previously published protocol paper. Phase I focused on data collection and involved search string generation, title and abstract screening, full-text review, construct assignment, and measure forward searches. During Phase II, we completed data extraction (i.e., coding psychometric information). Phase III involved data analysis, where two trained specialists independently rated each measurement tool using our psychometric rating criteria.

Results: Our team identified 124 measures of individual characteristics used in mental or behavioral health research, and 123 of those measures were deemed suitable for rating using Psychometric and Pragmatic Evidence Rating Scale. We identified measures of *knowledge and beliefs about the intervention* ($n = 76$), *self-efficacy* ($n = 24$), *individual stage of change* ($n = 2$), *individual identification with organization* ($n = 7$), and *other personal attributes* ($n = 15$). While psychometric information was unavailable and/or unreported for many measures, information about internal consistency and norms were the most commonly identified psychometric data across all individual characteristics' constructs. Ratings for all psychometric properties predominantly ranged from "poor" to "good."

Conclusion: The majority of research that develops, uses, or examines implementation measures that evaluate individual characteristics does not include the psychometric properties of those measures. The development and use of psychometric reporting standards could advance the use of valid and reliable tools within implementation research and practice, thereby enhancing the successful implementation and sustainment of evidence-based practice in community care.

Plain Language Summary: Measurement is the foundation for advancing practice in health care and other industries. In the field of implementation science, the state of measurement is only recently being targeted as an area for improvement, given that high-quality measures need to be identified and utilized in implementation work to avoid developing another research to practice gap. For the current study, we utilized the Consolidated Framework for Implementation Research

[1]Hathaway-Sycamores Child and Family Services, Pasadena, CA, USA
[2]Judge Baker Children's Center, Boston, MA, USA
[3]Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA
[4]Faculty of Health Sciences, Curtin University, Perth, WA, Australia
[5]Department of Global Health, University of Washington, Seattle, WA, USA

[6]Brown School, Washington University in St. Louis, St. Louis, MO, USA

Corresponding author:
Cameo Stanick, Hathaway-Sycamores Child and Family Services, 100 W. Walnut Street, Ste 375, Pasadena, CA 91124, USA.
Email: cstanick@hscfs.org

to identify measures related to individual characteristics' constructs, such as *knowledge and beliefs about the intervention, self-efficacy, individual identification with the organization, individual stage of change*, and *other personal attributes*. Our review showed that many measures exist for certain constructs (e.g., measures related to assessing providers' attitudes and perceptions about evidence-based practice interventions), while others have very few (e.g., an individual's stage of change). Also, we rated measures for their psychometric strength utilizing an anchored rating system and found that most measures assessing individual characteristics are in need of more research to establish their evidence of quality. It was also clear from our results that frequency of use/citations does not equate to high quality, psychometric strength. Ultimately, the state of the literature has demonstrated that assessing individual characteristics of implementation stakeholders is an area of strong interest in implementation work. It will be important for future research to focus on clearly delineating the psychometric properties of existing measures for saturated constructs, while for the others the emphasis should be on developing new, high-quality measures and make these available to stakeholders.

The field of implementation science aims to reduce the research-to-practice gap that impedes the delivery of evidence-based practices (EBPs) in routine care (Carnine, 1997). Numerous theories and frameworks have identified mechanisms and determinants known to influence the implementation of EPBs within community practice, including the characteristics of individuals who are inherently involved with the intervention and/or influence the implementation process (Albers et al., 2017; Moullin et al., 2015). Behavioral science has included studies of determinants of human action for decades and has developed long standing theories for examining behavior (e.g., theory of planned behavior [Ajzen, 1991]; transtheoretical model [Prochaska et al., 2001]). More recently, these concepts have been incorporated into multilevel theories and frameworks related to implementation (e.g., behavioral change wheel [Michie et al., 2011]). Such additions to implementation theories and frameworks have advanced implementation science by guiding high-quality qualitative studies. Notwithstanding the ongoing importance of qualitative studies, the ability to quantitatively measure individual characteristics is practical and would be advantageous to the field (e.g., measuring the differential impact of individual characteristics and inner setting on implementation outcomes, examining mechanisms of action of implementation strategies targeting implementers). Identification of psychometrically strong and pragmatic measures of individuals' characteristics is required to achieve such research advancements.

The Consolidated Framework for Implementation Research (CFIR) divides the characteristics of individuals that influence the implementation of EBPs into five primary constructs: knowledge and beliefs about the intervention, self-efficacy, individual stage of change, individual identification with the organization, and other personal attributes (Damschroder et al., 2009). Please reference Table 1 for definitions of these constructs. Ongoing research is attempting to understand how the interplay between individuals and the organizations in which they work uniquely impacts implementation processes and outcomes (e.g., Bearman et al., 2013; Brothers et al., 2015; Durlak & DuPre, 2008). For instance, a study by Eccles et al. (2011) utilized four measures to evaluate how the interaction between individual provider characteristics (e.g., self-reported cognitions about their organization and diabetes behaviors) and organizational factors (e.g., team climate) impacted providers' use of best practice diabetes management behaviors and interventions. Other researchers have evaluated how individual characteristics (e.g., clinician attitudes about EBPs and knowledge of cognitive-behavior therapy) impact other implementation outcomes, such as penetration and sustainment of EBPs for specific clinical problem areas (Edmunds et al., 2014). Such research, as well as advances in implementation practice, will remain limited unless stakeholders (e.g., researchers, implementation intermediaries, community practitioners) are able to use psychometrically strong and pragmatic measures to formally evaluate and monitor the impact of individual characteristics throughout various phases of implementation.

The identification and use of *psychometrically strong* implementation measures remains a significant barrier to advancing implementation research and practice (Halko et al., 2017; Rabin et al., 2010; Squires et al., 2011). For instance, Chaudoir and colleagues (2013) completed a literature review, extracting 125 full-text articles from various databases, to identify 112 measures designed to evaluate implementation constructs that can predict the implementation of EBPs. Thirty-five of these measures were related to individual characteristics, though very few of the measures identified within the review reported even basic psychometric properties (e.g., criterion validity). Additional scientific reviews have attempted to identify measures designed to evaluate individual characteristics (e.g., knowledge and beliefs, self-efficacy, individual

Table 1. Consolidated framework for implementation research (CFIR), characteristics of individuals construct definitions..

Construct	Definition
Knowledge and beliefs about the intervention	Individuals' attitudes toward and value placed on the intervention as well as familiarity with facts, truths, and principles related to the intervention.
Self-efficacy	Individual belief in their own capabilities to execute courses of action to achieve implementation goals.
Individual stage of change	Characterization of the phase an individual is in, as he or she progresses toward skilled, enthusiastic, and sustained use of the intervention.
Individual identification with organization	A broad construct related to how individuals perceive the organization, and their relationship and degree of commitment with that organization.
Other personal attributes	A broad construct to include other personal traits such as tolerance of ambiguity, intellectual ability, motivation, values, competence, capacity, and learning style.

CFIR: consolidated framework for implementation research.
Note: From http://www.cfirguide.org/constructs.html.

stage of change), but these reviews provided limited guidance for the selection of psychometrically strong measures because the reviews were specific to studies testing associations (Eccles et al., 2006), focused on particular provider types (e.g., nurses; Squires et al., 2011), were based on individual assessment items rather than implementation constructs (Chaudoir et al., 2013), or evaluated limited psychometric properties (e.g., criterion validity only; Chaudoir et al., 2013). These are significant issues, as well-developed measures should be evaluated and used in their entirety because their psychometric properties are related to all scales and subscales. The current review aimed to complement and advance upon previous work by identifying the implementation measures designed to assess individual characteristics within the current literature and evaluate those measures using a psychometric rating scale developed by Lewis and colleagues (2018). The results of the review may help stakeholders explore, select, and use high-quality measures to facilitate implementation research and practice in future implementation work.

Method

Design overview

The larger project from which this study derived was funded by the National Institute of Mental Health (NIMH) entitled, "Advancing implementation science through measure development and evaluation," and full details of the systematic review protocols have been published elsewhere (Lewis et al., 2018). For all projects, the systematic literature search consisted of three phases. Phase I was *measure identification*, which included the following five steps: (1) search string generation, (2) title and abstract screening, (3) full-text review, (4) measure assignment to characteristics of individuals and/or its subconstructs, and (5) measure forward (cited-by) searches. Phase II was *data extraction*, which consisted of coding relevant psychometric information. In Phase III, *data analysis* was completed.

Phase I: measure identification

PubMed and Embase bibliographic databases were used in the literature searches. Search strings were developed in consultation from PubMed support specialists and a library scientist. We utilized (1) terms for implementation (e.g., diffusion, knowledge translation, adoption); (2) terms for measurement (e.g., instrument, survey, questionnaire); (3) terms for evidence-based practice (e.g., innovation, guideline, empirically-supported treatment); and (4) terms for behavioral health (e.g., behavioral medicine, mental disease, psychiatry), which were consistent with our funding sources and our focus in behavioral health and related fields (Lewis et al., 2018). For the current study, we included a fifth level for each of the following *Characteristics of Individuals* constructs from the CFIR (Damschroder et al., 2009): (1) *knowledge and beliefs*, (2) *self-efficacy*, (3) *individual stage of change*, (4) *individual identification with the organization*, and (5) *other personal attributes*. Literature searches were conducted independently for each construct; thus, five different sets of search strings were employed. The time frame for articles published was from 1985 to 2017, and searches were completed from April 2017 to May 2017.

For inclusion criteria, two trained research specialists (CD, KM) identified articles through a title and abstract screening, followed by full-text review, to confirm relevance to the study parameters. Empirical studies that contained one or more quantitative measures of any of the five CFIR constructs were included if they were used in an evaluation of an implementation effort in a behavioral health context. Measures were excluded if they were considered "unsuitable for rating" based on the format of measure if it did not produce psychometric information (e.g., qualitative nomination form) or format did not lend itself to the rating system described below (e.g., cost analysis formula, penetration formula). See Appendix 1 for the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagrams, providing a breakdown of the inclusion/exclusion criteria.

Trained research specialists (CD, KM) then completed the fourth step in which they used a consensus coding

approach to assigned included measures to one or more of the five CFIR constructs (Bradley et al., 2007; Damschroder et al., 2009). In our consensus coding approach, the raters independently coded measures into one of the five CFIR constructs and met approximately weekly to discuss any disagreements. When disagreements occurred, they were discussed, analyzed, and further discussed until consensus was found. We used the study author's definition of which construct was being measured; however, in the absence of a definition, trained research specialists completed content coding (Lewis et al., 2018). Content experts (CS, HH) also reviewed each item within each measure and/or scale to confirm or reassign as necessary. Finally, measures were subjected to "cited-by" searches in PubMed and Embase to identify all empirical literature that included the measure in a behavioral health context.

Phase II: data extraction

Following the "cited-by" searches, all relevant literature was compiled into "measure packets" that included the measure itself (as available), the original measurement development article(s) (or article with the first empirical use in a behavioral health context), and all additional empirical articles describing the uses of the measure in behavioral health. Trained research specialists (CD, KM) reviewed each article and electronically extracted all information relevant to the psychometric and pragmatic rating criteria, referred to hereafter as PAPERS (Psychometric and Pragmatic Evidence Rating Scale). The full rating system and criteria for the PAPERS are published elsewhere (Stanick et al., 2019; Lewis et al., 2018). The current study focuses on the nine psychometric criteria identified in the PAPERS system: (1) internal consistency, (2) convergent validity, (3) discriminant validity, (4) known-groups validity, (5) predictive validity, (6) concurrent validity, (7) structural validity, (8) responsiveness, and (9) norms. These psychometric criteria were extracted and assessed for both full measure and individual scale/subscale levels as appropriate.

The PAPERS criteria have an assigned anchor scale, which ranges from, "poor" (-1), "none" (0), "minimal/emerging" (1), "adequate" (2), "good" (3), to "excellent" (4). Each anchor has an accompanying operationalized definition relevant to each criterion. To calculate a final score, either a single score was identified (if a measure had only one rating for a criterion) *or* we utilized a "rolled up median" approach. If a measure was used in multiple studies and the same criterion was reported in multiple studies, we calculated a median score across the relevant articles to generate the final rating of measure on that criterion. This was completed for all criteria referenced for each measure. Furthermore, if a measure contained a set of scales relevant to a construct, the ratings for those individual (sub)

scales were "rolled up" by calculating a median score which was then used as the final aggregate rating for the whole measure on the relevant criterion. For instance, if a measure had three scales that were dimensions of *self-efficacy*, and each had a rating for internal consistency (because this psychometric criterion was reported in the relevant articles), the median of those ratings was calculated and assigned as the final rating of internal consistency for the full measure. In the case that the calculated median resulted in a non-integer rating, the non-integer was rounded down (e.g., internal consistency ratings of 2 and 3 would result in a 2.5 median which was rounded down to 2), and in the case that the median of two scores was "0" (e.g., a score of -1 and 1) the lower score was used (e.g., -1). Other characteristics, such as descriptive data, were gathered and reported on each measure as well (i.e., country of origin; concept defined by authors; number of articles contained in each measure packet; number of scales; number of items; setting in which measure had been used; level of analysis; target problem; and stage of implementation as defined by the Exploration, Adoption/Preparation, Implementation, Sustainment (EPIS) model [Aarons et al., 2011]).

Phase III: data analysis

For each measure, a total score was calculated by summing each of the rating criterion scores. Simple statistics (e.g., frequencies) were calculated on relevant psychometric data. Possible total score calculations could range from -9 (i.e., each rating criterion was equal to -1) to 36 (i.e., each rating criterion was equal to the highest possible score of 4). To be able to visually examine how measures compared to one another in their ratings, bar graphs were generated that display head-to-head comparisons of total scores for all measures, across all criteria, within a given construct. These are shown in Figures 1 to 5.

Results

Of the 223 total measures of characteristics of individuals (including subscales), only one was categorized as unsuitable for rating the psychometric evidence. The majority of results described below are presented at the level of whole measures. Whenever appropriate we include the number of subscales relevant to a construct within that measure.

Overview of measures

Across the five subconstructs related to *characteristics of individuals* (i.e., *knowledge and beliefs, self-efficacy, individual stage of change, individual identification with an organization, other personal attributes*), 112 measures were used in mental or behavioral health care research and
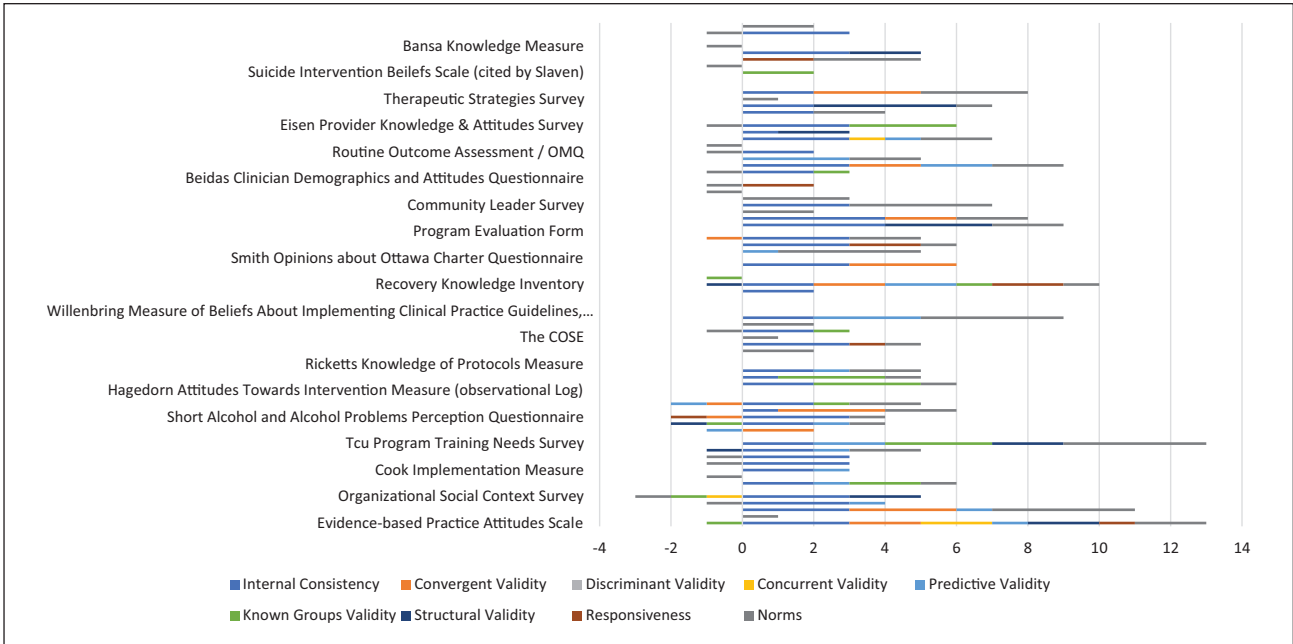
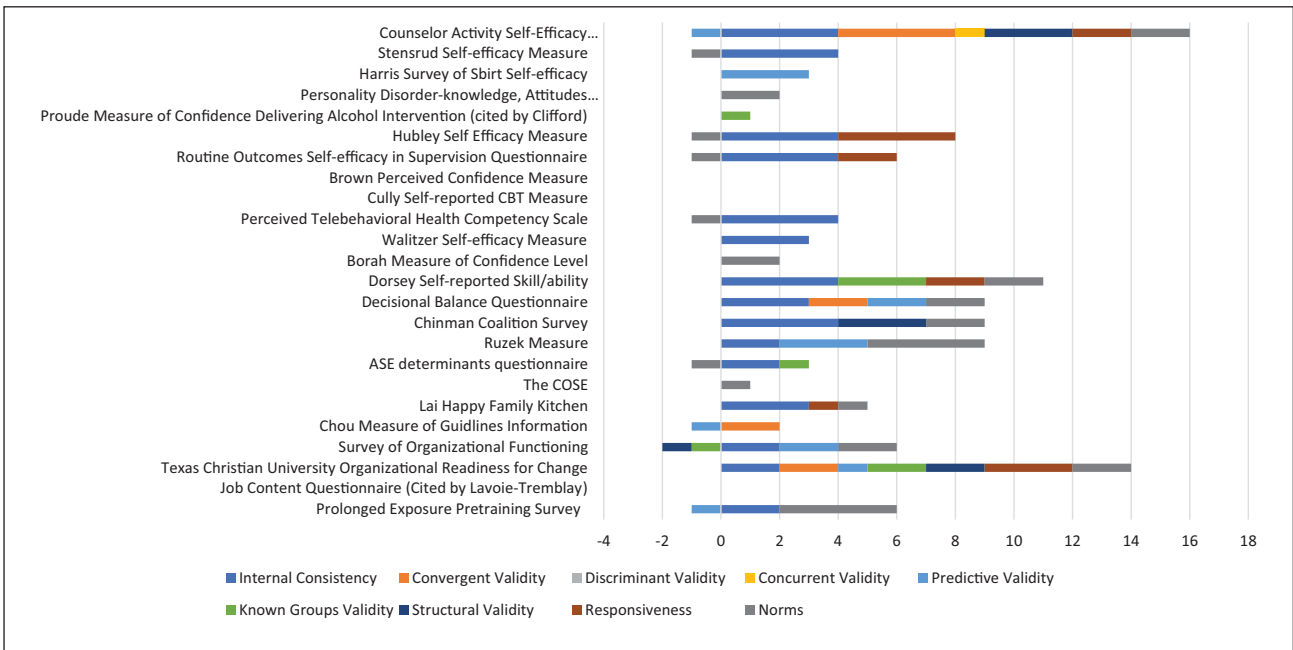**Figure 1.** Knowledge and beliefs about the intervention.

Figure 2. Self-efficacy.

identified across our electronic database searches. There were 104 measures of *knowledge and beliefs about the intervention*, and one was not suitable for rating (measures were considered "not suitable for rating" if the format of construct assessment did not produce psychometric information or format of the measure did not conform to the rating scale). Twenty-eight subscales were scored within the *knowledge and beliefs about the intervention* construct (e.g., ASE Determinants Questionnaire—Knowledge and Skills concerning the use of the guidelines for depression subscale; Zwerver et al., 2013). Twenty-four measures of *self-efficacy* were identified, with 16 additional subscales
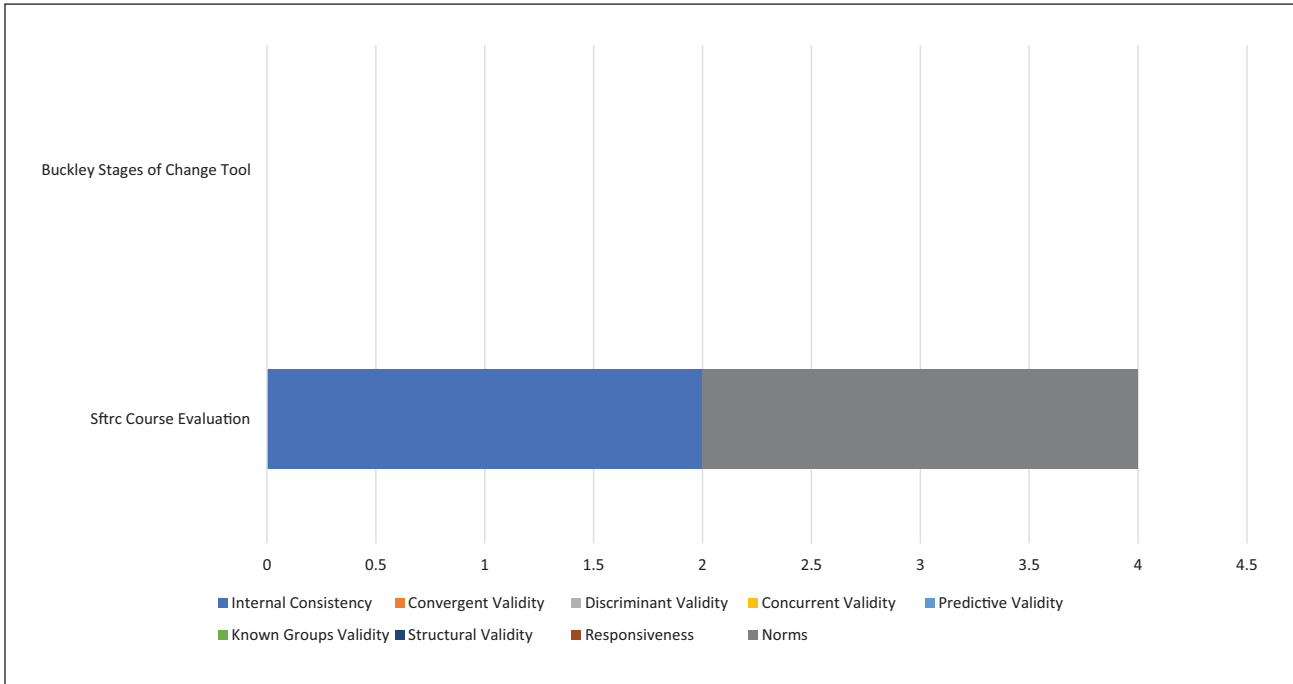
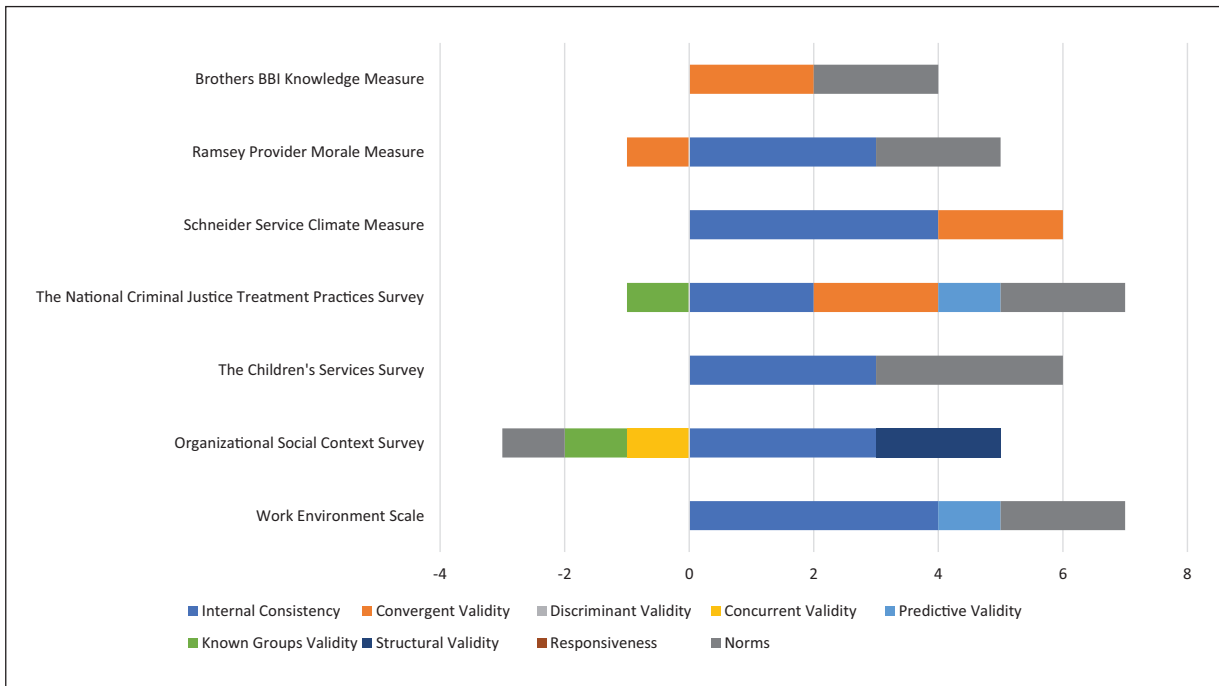**Figure 3.** Individual stage of change.

Figure 4. Individual identification with the organization.

identified. Two measures of *individual stage of change* were identified. Seven measures of *individual identification with the organization* were identified, as well as three subscales from "parent" measures that specifically

assessed an individual's alignment with their organization (e.g., The National Criminal Justice Treatment Practices Survey—Organizational Commitment sub-scale; Taxman et al., 2007). Fifteen measures of *other personal attributes*
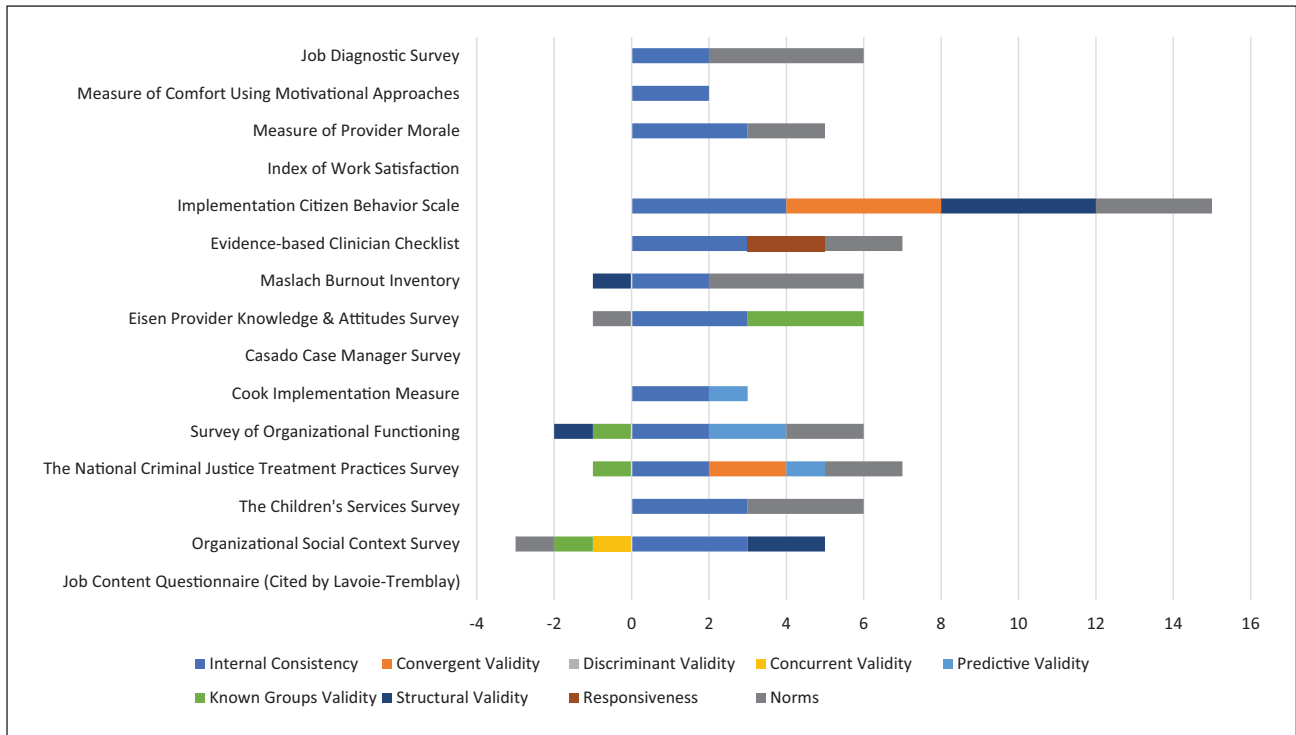
Figure 5. Other personal attributes.

were identified, with 41 additional subscales pertinent to *other personal attributes* (e.g., burnout, job frustration, capability).

Characteristics of measures

The descriptive characteristics of measures can be found in Table 2. Of the 112 full measures (no subscales) that were suitable for rating, over half were single-use only ($n=75$; 67%). Most were created in the United States ($n=80$; 71%), whereas the remaining measures were developed in several different countries, including Australia, Canada, Catalonia, China, the United Kingdom, the Netherlands, Sweden, Finland, Italy, Korea, Norway, South Africa, Poland, and Japan. When it could be determined, many measures were applied in contexts outside of an identified behavioral health setting (e.g., university, primary care, hospitals, etc.; $n=50$, 45%). An additional 30% ($n=33$) were deployed in an outpatient community health context. Finally, approximately one-third ($n=35$, 31%) of measures identified could be used to assess characteristics of individuals that influence implementation for specific purposes (e.g., psychopathologies, children's mental health, suicide prevention), and 23% ($n=26$) were specifically used in the context of targeting substance abuse. For most measures, we were unable to determine in which stage of implementation (exploration, planning, implementation, sustainability) the measures were applied due to limitations in reporting ($n=107$; 96%).

Availability of psychometric evidence

Of the 223 total measures of characteristics of individuals (including subscales), only one was categorized as unsuitable for rating the psychometric evidence (i.e., from the *knowledge and beliefs about the intervention* subconstruct). The remaining measures had varying degrees of psychometric information (Table 3). Approximately one-third of measures had no evidence of internal consistency ($n=79$, 36%). One hundred seventy nine measures (81%) had no evidence of convergent validity, 216 (97%) had no evidence of concurrent validity, 176 (79%) had no evidence of predictive validity, 187 (84%) had no evidence of known-groups validity, 197 (89%) had no evidence of structural validity, and 195 (88%) had no evidence of responsiveness. Finally, no measures had evidence of discriminant validity.

Psychometric evidence rating scale results

Table 4 describes the psychometric evidence available for measures for which information was available (e.g., those with non-zero ratings on the PAPERS criteria; $n=86$). Median ratings and range of ratings for psychometric properties are provided.

Knowledge and beliefs about the intervention. One-hundred four measures were identified in regard to *knowledge and beliefs about the intervention* and were used in a mental or

Table 2. Description of measures and subscales.

	Knowledge & beliefs (n = 79)		Self-efficacy (n = 24)		Individual stage of change (n = 2)		Individual identification with organization (n = 7)		Other personal attributes (n = 15)	
	n	%	n	%	n	%	n	%	n	%
Concept defined										
Yes	18	23	3	13	0	0	3	43	7	47
No	61	77	21	88	2	100	4	57	8	53
One-time use only										
Yes	49	62	17	71	2	10	3	43	9	60
No	30	38	7	29	0	0	4	57	6	40
Number of scales[a]										
1	0	0	0	0	0	0	0	0	0	0
2 to 5	16	20	11	46	0	0	0	0	2	13
6 or more	4	5	5	21	0	0	5	71	2	13
Number of items[b]										
1 to 5	0	0	2	8	0	0	0	0	0	0
6 to 10	3	4	3	13	0	0	2	29	2	13
11 or more	32	42	16	67	2	100	3	43	5	33
Country										
USA	32	42	16	67	1	50	7	100	6	40
Other	17	22	8	33	1	50	0	0	9	60
Setting[c]										
Inpatient psychiatry	0	0	0	0	0	0	0	0	1	7
Outpatient community	16	20	7	29	0	0	2	29	4	27
School mental health	0	0	1	4	0	0	0	0	0	0
Residential care	1	1	0	0	0	0	0	0	1	7
Other	27	34	13	54	1	50	1	14	3	20
Level[c]										
Provider	42	53	17	71	0	0	7	100	5	33
Supervisor	0	0	1	4	0	0	0	0	1	7
Other	4	5	0	0	1	50	0	0	1	7
Population[c]										
General mental health	2	3	1	4	0	0	0	0	0	0
Anxiety	1	12	0	0	0	0	0	0	0	0
Depression	3	34	0	0	0	0	0	0	0	0
Alcohol use disorder	1	12	1	4	0	0	0	0	0	0
Substance use disorder	11	14	5	21	1	50	2	29	2	29
Behavioral disorder	1	1	0	0	0	0	0	0	0	0
Mania	3	4	0	0	0	0	0	0	0	0
Trauma	3	4	2	8	0	0	0	0	0	0
Other	18	23	9	38	0	0	3	43	2	29

[a]Some measures did not report total number of scales.
[b]Some measures did not report total number of items.
[c]Some measures used in multiple settings, levels, and populations.

behavioral health context. Evidence of psychometric strength was limited. Specifically, only 39 measures had evidence for internal consistency (37.5%), 17 had evidence of predictive validity (16%), 14 had evidence of known-groups validity (13.5%), 12 had evidence of convergent validity (11.5%), 10 measures had evidence of structural validity, (10%), seven had evidence of responsiveness (6.7%), and three had evidence of concurrent validity (2.8%). Seventy-nine measures had non-zero ratings for norms (76%), and no measures had evidence of discriminant validity.

For those measures with information available (e.g., those with non-zero ratings), the median rating for internal consistency was "2—adequate," "2—adequate" for

Table 3. Psychometric information availability (with subscales).

	Knowledge & beliefs (n = 104)		Self-efficacy (n = 40)		Individual stage of change (n = 2)		Individual identification with organization (n = 10)		Other personal attributes (n = 56)	
	n	%	n	%	n	%	n	%	n	%
Internal consistency	50	48	28	70	1	50	8	80	46	82
Convergent validity	14	14	11	28	0	0	5	50	7	13
Discriminant validity	0	0	0	0	0	0	0	0	0	0
Known-groups validity	17	16	7	18	0	0	2	20	7	13
Predictive validity	19	18	10	25	0	0	4	40	10	18
Concurrent validity	3	3	1	3	0	0	1	10	1	2
Structural validity	11	11	5	13	0	0	2	20	5	9
Responsiveness	9	9	13	33	0	0	0	0	2	4
Norms	65	63	30	75	1	50	8	80	47	84

Table 4. Summary statistics for instrument ratings.

	Knowledge & beliefs (n = 76)		Self-efficacy (n = 24)		Individual stage of change (n = 2)		Individual identification with organization (n = 7)		Other personal attributes (n = 15)	
	Mdn	R	Mdn	R	Mdn	R	Mdn	R	Mdn	R
Internal consistency	2	1.4	3	2.4	2	–	3	2.4	2	2.4
Convergent validity	2	−1.3	2	2.4	–	–	2	−1.2	3	2.4
Discriminant validity	–	–	–	–	–	–	–	–	–	–
Known-groups validity	1	−1.3	1	−1.3	–	–	−1	–	−1	−1.3
Predictive validity	1	−1.3	1	−1.3	–	–	1	–	1	1.2
Concurrent validity	1	−1.1	1	–	–	–	−1	–	−1	–
Structural validity	2	−1.4	2	−1.3	–	–	2	–	0	−1.4
Responsiveness	2	−1.2	2	1.4	–	–	–	–	2	–
Norms	1	−1.4	2	−1.4	2	–	2	−1.3	2	−1.4

Note: *Mdn*: median, excluding zeros where psychometric information not available and measures that were deemed unsuitable for rating. R: range.

convergent validity, "−1—poor" for concurrent validity, "−1—poor" for predictive validity, "−1—poor" for known-groups validity, "2—adequate" for structural validity, "−1—poor" for responsiveness, and "−1—poor" for norms.

The Texas Christian University Training Needs Survey received the highest psychometric rating total score among all measures of *knowledge and beliefs* (total maximum score = 13; maximum possible score = 36), with ratings of "2—adequate" for internal consistency, "2—adequate" for predictive validity, "3—good" for known-groups validity, "2—adequate" for structural validity, and "4—excellent" for norms (Simpson, 2002).

Self-efficacy. Twenty-four measures designed to assess *self-efficacy* were identified within mental or behavioral health care research. Evidence about internal consistency was available for 15 measures (63%), convergent validity for four measures (17%), discriminant validity for no measures (0%), concurrent validity for one measure (4%), predictive validity for eight measures (33%), known-groups validity for five measures (21%), structural validity

for four measures (17%), responsiveness for six measures (25%), and norms for 17 measures (71%). For all measures of *self-efficacy* that reported data for the PAPERS criteria (i.e., those with non-zero ratings), the median rating for internal consistency was "3—good," "2—adequate" for convergent validity, "1—minimal/emerging" for concurrent validity, "1—minimal/emerging" for predictive validity, "1—minimal/emerging" for known-groups validity, "2—adequate" for structural validity, "2—adequate" for responsiveness, and "2—adequate" for norms.

The Counselor Activity Self-Efficacy Scales (CASES) received the highest psychometric rating score in comparison to all measures of *self-efficacy* found within mental and behavioral health care research publications. The psychometric total maximum score for the scale equaled 15 (maximum possible score = 36). The scale obtained ratings of "4—excellent" for internal consistency, "4—excellent" for convergent validity, "1—minimal/emerging" for concurrent validity, "−1—poor" for predictive validity, "3—good" for structural validity, "2—adequate" for responsiveness, and "2—adequate" for norms (Lent et al., 2003). There was no information available for the

remaining PAPERS psychometric criteria. The PAPERS scores provided for the Counselor Activity Self-Efficacy scale were calculated based on 17 uses of this measure in behavioral health care research.

Individual stage of change. Two measures designed to assess *individual stage of change* were identified within mental or behavioral health care research. Only one of the identified measures, the San Francisco Treatment Research Center Course Evaluation (Haug et al., 2008), provided any information about psychometric properties; therefore, median scores across measures will not be reported. The SFTRC Course Evaluation received a psychometric total maximum score equaling 15 (maximum possible score = 36). The scale obtained ratings of "2—adequate" for internal consistency and "2—adequate" for norms. There was no information available for the remaining PAPERS psychometric criteria. The PAPERS scores provided for the SFTRC Course Evaluation were calculated based on one use of this measure in behavioral health care research.

Individual identification with the organization. Seven measures designed to assess *individual identification with the organization* were identified within mental or behavioral health care research. Evidence about internal consistency was available for six measures (86%), convergent validity for four measures (57%), discriminant validity for no measures (0%), concurrent validity for one measure (14%), predictive validity for two measures (29%), known-groups validity for two measures (29%), structural validity for one measure (14%), responsiveness for no measures (0%), and norms for six measures (86%). For all measures of *individual identification with the organization* that reported data for the PAPERS criteria (i.e., those with non-zero ratings), the median rating for internal consistency was "3—good," "2—adequate" for convergent validity, "−1—poor" for concurrent validity, "1—minimal/emerging" for predictive validity, "−1—poor" for known-groups validity, "2—adequate" for structural validity, and "2—adequate" for norms.

The Work Environment Scale received the highest psychometric rating score in comparison to all measures of *individual identification with the organization* found within mental and behavioral health care research publications. The psychometric total maximum score for the scale equaled 7 (maximum possible score = 36). The scale obtained ratings of "4—excellent" for internal consistency, "1—minimal/emerging" for predictive validity, and "2—adequate" for norms (Moos & Insel, 1974). There was no information available for the remaining PAPERS psychometric criteria. The PAPERS scores provided for the Work Environment Scale were calculated based on six uses of this measure in behavioral health care research.

Other personal attributes. Fifteen measures of *other personal attributes* were identified within the context of mental or behavioral health care research contexts. Examples of *other personal attributes* included characteristics such as autonomy in one's job, one's experience of burnout, and role clarity. Psychometric evidence was available for 12 measures for internal consistency (80%), 10 measures had evidence of norms (67%), four measures for known-groups validity (27%), four measures of structural validity (27%), three measures of predictive validity (20%), two measures for convergent validity (13%), and one measure each with evidence of concurrent validity and responsiveness (7%). No measures had evidence of discriminant validity.

For those measures of *other personal attributes* with information available for rating (i.e., those with non-zero ratings on PAPERS criteria), the median rating for internal consistency was "2—adequate," for convergent validity it was "3—good," "1—minimal/emerging" for predictive validity, "−1—poor" for known-groups validity, "2—adequate" for responsiveness, and "2—adequate" for norms. Concurrent validity only had one non-zero rating (−1) and therefore the median was "−1—poor." For structural validity, the median was 0.5 and using the worst score counts was rounded to zero.

The Implementation Citizen Behavior Scale had the highest psychometric rating total score among measures of *other personal attributes* (psychometric total maximum score = 15; maximum possible score = 36; Ehrhart et al., 2015). This measure received scores of "4—excellent" for internal consistency, convergent validity, and structural validity, and a score of "3—good" for the norms criterion. No information was available on the remaining psychometric criteria.

Discussion

Characteristics of individuals and many of the subconstructs therein—particularly *knowledge and beliefs about the intervention*—have a high number of associated measures, which suggests that implementation leaders and researchers recognize the potentially strong influence of implementation stakeholders. Indeed, this particular construct appears to have had consistent focus in implementation efforts, dating back to Rogers' *Diffusion of Innovations* and the role of stakeholder attitudes toward innovations (Rogers, 1995). Measures have been developed to assess stakeholder knowledge, beliefs, and perceptions of EBPs, which have commonly been examined as intervention targets and predictors of adoption. The strong focus on knowledge and beliefs might have contributed to researchers ignoring other factors that may have as much as (or more) influence on implementation successes or failures (e.g., self-efficacy and individual stage of change). Only recently have researchers expanded their perspectives and

analyses to include the other four domains of *characteristics of individuals* (Brothers et al., 2015).

As with other areas of implementation measurement, the quality of existing measures assessing the characteristics of individuals is lacking (Lewis et al., 2015). Importantly, quantity of literature or measures speaks little to the quality of psychometric data available—Even the highest scoring measure in the *knowledge and beliefs* construct was far below the maximum score (13 out of a maximum of 36; Weiner et al., 2020). These results suggest that the need to systematically develop and evaluate the psychometric properties is warranted for this implementation domain and the majority of its constructs. Although evidence of internal consistency and norms was available for most measures (scored at the "good" level across all measures), evidence of other psychometric criteria (e.g., structural validity, known-groups validity, predictive validity, etc.) was lacking. Notably, the ability for a measure of individual characteristics to meaningfully detect change in an implementation outcome (e.g., fidelity) or correlate as hypothesized with another construct of interest (e.g., readiness for implementation) seems paramount as existing research suggests this relationship is critical and yet less than 20% of measures had any evidence of predictive validity testing/indicators (Aarons et al., 2012). Given the number of available measures and the high volume of research for select measures (e.g., EBPAS-50; Aarons et al., 2010), it was surprising that *responsiveness*, information about a measure's sensitivity to change, was only available for 12.5% of all measures assessed. Similarly, only 17% of measures had assessed structural validity, even though examining the dimensionality of measures is pertinent even before assessing its internal consistency (DeVellis, 2012).

The reasons for such a paucity of psychometric evidence are likely many. For one, journals typically require certain features, such as internal consistency, in reporting standards but do not necessarily require other psychometric properties (Weiner et al., 2020). Furthermore, Lewis et al. (2015) offered suggestions for why reporting on psychometric properties is low, which include the notion that many implementation measures may be developed in the context of a single study for idiosyncratic needs of specific projects without the contributions of a psychometrician. Indeed, our results revealed that 67% of measures within the *characteristics of individuals* domain were "single-use," suggesting that a substantial portion may have been developed for specific, immediate projects.

The measures for highest overall rating of psychometric properties were the Counselor Activity Self-Efficacy Scale (Lent et al., 2003) for the *self-efficacy* construct and the Implementation Citizen Behavior Scale (Ehrhart et al., 2015) for the *other personal attributes* construct. For both measures, internal consistency and convergent validity were rated in the "excellent" range. The Implementation

Citizen Behavior Scale was also rated "excellent" for structural validity. Notably, these scores resulted from relatively few uses (i.e., 1 use of the Implementation Citizen Behavior Scale and 17 uses of the Counselor Activity Self-Efficacy Scale). Ultimately, the field of implementation science, and development of implementation specific measures, is still young and measures require repeated usage *and* targeted reporting on psychometric characteristics to establish their evidence. As found in Weiner et al. (2020), frequency of use was not necessarily correlated with higher psychometric rating scores. For example, the Texas Christian University Organizational Readiness for Change (TCU-ORC) measure scored 14 out of 36 and had the third highest score. However, it was most notable that the measure is widely used in the implementation science literature (cited 68 + times) and yet psychometric ratings have not improved substantially with repeated use (Lehman et al., 2002).

Limitations

The current study had limitations. Specifically, this systematic review and rating process was focused entirely on measures pertaining to *characteristics of individuals* and its constructs utilized in mental or behavioral health contexts. It is possible that measures assessing this domain exist in other contexts. For instance, some measures identified in the current review were related to characteristics of individuals within their job contexts in general (e.g., Job Diagnostic Survey and Job Content Questionnaire), which suggests that other fields such as business, industry, and so on, may use measures that were beyond the scope of our review (Hackman & Oldham, 1975; Karasek, 1985). Along these lines, it is also possible that some of the measures of *characteristics of individuals* were utilized outside of mental or behavioral health contexts, which means that we may have missed these opportunities for rating psychometric properties. In the event a source/measure development article was available, however, we always attempted to locate and rate the psychometric criteria within regardless to reduce the likelihood that psychometric properties would be left unrated. Furthermore, we conducted our measures-forward searches through 2017, and it is possible that if a measure was developed between 2017 and the timing of this publication that it would not have been included in our review. In addition, the current study focused on applying the psychometric aspects of the PAPERS criteria to measures of *characteristics of individuals.* As previously referenced, the pragmatic aspects of the PAPERS criteria have been published elsewhere (Stanick et al., et al., 2019; Lewis et al., 2018) and their application across the 45 + implementation science measures is still in development.

Furthermore, the reporting requirements of journals may impact available features for rating with the PAPERS

criteria. For instance, journal requirements may include listing certain statistics, such as internal consistency, but not other psychometrics. Also, some criteria, such as structural validity, may have been addressed via factor analysis but authors may have failed to report the amount of variance explained by different factors and/or report key model statistics. The extent and quality of reporting features included in articles ultimately influenced our ability to accurately rate measures using PAPERS criteria.

Conclusions

In total, there are many measures across multiple constructs assessing various *characteristics of individuals* involved in implementation. Certain constructs have a plethora of available measures assessing similar features (e.g., *knowledge and beliefs about the intervention*), whereas others have very few, and future research could be concentrated in developing psychometrically strong measures in these areas (e.g., *individual stage of change, individual identification with organization*). For those constructs where many measures exist and/or where certain measures have been utilized many times, researchers and implementation leaders alike would benefit from more focused, concentrated efforts to delineate the psychometric properties of those measures and ensure that high-quality measures are prioritized for use in implementation work.

Declaration of conflicting interests

Funding

ORCID iDs

Cameo Stanick https://orcid.org/0000-0002-6076-3726
Kayne Mettert https://orcid.org/0000-0003-1750-7863
Byron Powell https://orcid.org/0000-0001-5245-1186
Cara C Lewis https://orcid.org/0000-0001-8920-8075

References

Aarons, G. A., Glisson, C., Hoagwood, K., Kelleher, K., Landsverk, J., & Cafri, G. (2010). Psychometric properties and U.S. National norms of the Evidence-Based Practice Attitude Scale (EBPAS). *Psychological Assessment*, *22*(2), 356–365. https://doi.org/10.1037/a0019188

Aarons, G. A., Green, A. E., & Miller, E. (2012). Researching readiness for implementation of evidence-based practice: A comprehensive review of the Evidence-Based Practice Attitude Scale (EBPAS). In B. Kelly & D. F. Perkins (Eds.), *Handbook of implementation science for psychology in education* (pp. 150–164). Cambridge University Press.

Aarons, G. A., Hurlburt, M., & Horwitz, S. M. (2011). Advancing a conceptual model of evidence-based practice implementation in public service sectors. *Administration and Policy in Mental Health*, *38*(1), 4–23. https://doi.org/10.1007/s10488-010-0327-7

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-t

Albers, B., Mildon, R., Lyon, A. R., & Shlonsky, A. (2017). Implementation frameworks in child, youth, and family services—Results from a scoping review. *Children and Youth Services Review*, *81*, 101–116. https://doi.org/10.1016/j.childyouth.2017.07.003

Bearman, S. K., Weisz, J. R., Chorpita, B. F., Hoagwood, K., Ward, A., Ugueto, A. M., Bernstein, A., & The Research Network on Youth Mental. (2013). More practice, less preach? The role of supervision processes and therapist characteristics in EBP implementation. *Administration and Policy in Mental Health*, *40*(6), 518–529. https://doi.org/10.1007/s10488-013-0485-5

Bradley, E. H., Curry, L. A., & Devers, K. J. (2007). Qualitative data analysis for health services research: Developing taxonomy, themes, and theory. *Health Services Research*, *42*(4), 1758–1772. https://doi.org/10.1111/j.1475-6773.2006.00684.x

Brothers, B. M., Carpenter, K. M., Shelby, R. A., Thornton, L. M., Frierson, G. M., Patterson, K. L., & Andersen, B. L. (2015). Dissemination of an evidence-based treatment for cancer patients: Training is the necessary first step. *Translational Behavioral Medicine*, *5*(1), 103–112. https://doi.org/10.1007/s13142-014-0273-0

Carnine, D. (1997). Bridging the research-to-practice gap. *Exceptional Children*, *63*(4), 513–521.

Chaudoir, S. R., Dugan, A. G., & Barr, C. H. (2013). Measuring factors affecting implementation of health innovations: A systematic review of structural, organizational, provider, patient, and innovation level measures. *Implementation Science*, *8*, 22. https://doi.org/10.1186/1748-5908-8-22

Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S. R., Alexander, J. A., & Lowery, J. C. (2009). Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. *Implementation Science*, *4*, 50. https://doi.org/10.1186/1748-5908-4-50

DeVellis, R. F. (2012). *Scale development: Theory and applications*. SAGE.

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *The American Journal of Community Psychology*, *41*(3–4), 327–350. https://doi.org/10.1007/s10464-008-9165-0

Eccles, M. P., Hrisos, S., Francis, J., Kaner, E. F., Dickinson, H. O., Beyer, F., & Johnston, M. (2006). Do self-reported intentions predict clinicians' behaviour: A systematic review.

Implementation Science, *1*, 28. https://doi.org/10.1186/1748-5908-1-28

Eccles, M. P., Hrisos, S., Francis, J. J., Stamp, E., Johnston, M., Hawthorne, G., Steen, N., Grimshaw, J., Elovainio, M., Presseau, J., & Hunter, M. (2011). Instrument development, data collection, and characteristics of practices, staff, and measures in the Improving Quality of Care in Diabetes (iQuaD) Study. *Implementation Science*, *6*, 61. https://doi.org/10.1186/1748-5908-6-61

Edmunds, J. M., Read, K. L., Ringle, V. A., Brodman, D. M., Kendall, P. C., & Beidas, R. S. (2014). Sustaining clinician penetration, attitudes and knowledge in cognitive-behavioral therapy for youth anxiety. *Implementation Science*, *9*, 89. https://doi.org/10.1186/s13012-014-0089-9

Ehrhart, M. G., Aarons, G. A., & Farahnak, L. R. (2015). Going above and beyond for implementation: The development and validity testing of the Implementation Citizenship Behavior Scale (ICBS). *Implementation Science*, *10*, 65. https://doi.org/10.1186/s13012-015-0255-8

Hackman, J. R., & Oldham, G. R. (1975). Development of the job diagnostic survey. *Journal of Applied Psychology*, *60*(2), 159–170. https://doi.org/10.1037/h0076546

Halko, H. M., Stanick, C. F., Dorsey, C. N., Weiner, B. J., Powell, B. J., & Lewis, C. C. (2017, December 4–6). *Stakeholder perspectives and use of implementation science measurement tools* [Conference session]. Paper presented at the 10th Annual Conference on the Science of Dissemination and Implementation, Arlington, VA, United States.

Haug, N. A., Shopshire, M., Tajima, B., Gruber, V., & Guydish, J. (2008). Adoption of evidence-based practices among substance abuse treatment providers. *Journal of Drug Education*, *38*(2), 181–192. https://doi.org/10.2190/DE.38.2.f

Karasek, R. (1985). *Job Content Questionnaire user's guide*. Department of Work Environment.

Lehman, W. E., Greener, J. M., & Simpson, D. D. (2002). Assessing organizational readiness for change. *Journal of Substance Abuse Treatment*, *22*(4), 197–209. https://doi.org/10.1016/s0740-5472(02)00233-7

Lent, R. W., Hill, C. E., & Hoffman, M. A. (2003). Development and validation of the Counselor Activity Self-Efficacy Scales. *Journal of Counseling Psychology*, *50*(1), 97–108. https://doi.org/10.1037/0022-0167.50.1.97

Lewis, C. C., Fischer, S., Weiner, B. J., Stanick, C., Kim, M., & Martinez, R. G. (2015). Outcomes for implementation science: An enhanced systematic review of instruments using evidence-based rating criteria. *Implementation Science*, *10*, 155. https://doi.org/10.1186/s13012-015-0342-x

Lewis, C. C., Mettert, K. D., Dorsey, C. N., Martinez, R. G., Weiner, B. J., Nolen, E., Stanick, C., Halko, H., & Powell, B. J. (2018). An updated protocol for a systematic review of implementation-related measures. *Systematic Reviews*, *7*(1), 66. https://doi.org/10.1186/s13643-018-0728-3

Michie, S., van Stralen, M. M., & West, R. (2011). The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science*, *6*, 42. https://doi.org/10.1186/1748-5908-6-42

Moos, R. H., & Insel, P. N. (1974). *Work Environment Scale*. Consulting Psychologists Press.

Moullin, J. C., Sabater-Hernandez, D., Fernandez-Llimos, F., & Benrimoj, S. I. (2015). A systematic review of implementation frameworks of innovations in healthcare and resulting generic implementation framework. *Health Research Policy and Systems*, *13*, 16. https://doi.org/10.1186/s12961-015-0005-z

Prochaska, J. M., Prochaska, J. O., & Levesque, D. A. (2001). A transtheoretical approach to changing organizations. *Administration and Policy in Mental Health*, *28*(4), 247–261. https://doi.org/10.1023/a:1011155212811

Rabin, B. A., Glasgow, R. E., Kerner, J. F., Klump, M. P., & Brownson, R. C. (2010). Dissemination and implementation research on community-based cancer prevention: A systematic review. *American Journal of Preventative Medicine*, *38*(4), 443–456. https://doi.org/10.1016/j.amepre.2009.12.035

Rogers, E. M. (1995). *Diffusion of innovations*. Free Press.

Simpson, D. D. (2002). A conceptual framework for transferring research to practice. *Journal of Substance Abuse Treatment*, *22*(4), 171–182. https://doi.org/10.1016/s0740-5472(02)00231-3

Squires, J. E., Estabrooks, C. A., Gustavsson, P., & Wallin, L. (2011). Individual determinants of research utilization by nurses: A systematic review update. *Implementation Science*, *6*, 1. https://doi.org/10.1186/1748-5908-6-1

Stanick, C. F., Halko, H. M., Nolen, E. A., Powell, B. J., Dorsey, C. N., Mettert, K. D., Weiner, B., Barwick, M., Wolfenden, L., Damschroder, L., & Lewis, C. C. (2019). Pragmatic measures for implementation research: Development of the Psychometric and Pragmatic Evidence Rating Scale (PAPERS). *Translational Behavioral Medicine*, *11*(1), 11–20. https://doi.org/10.1093/tbm/ibz164

Taxman, F. S., Young, D. W., Wiersema, B., Rhodes, A., & Mitchell, S. (2007). The National Criminal Justice Treatment Practices survey: Multilevel survey methods and procedures. *Journal of Substance Abuse Treatment*, *32*(3), 225–238. https://doi.org/10.1016/j.jsat.2007.01.002

Weiner, B., Mettert, K. D., Dorsey, C. N., Nolen, E. A., Stanick, C., Powell, B. J., & Lewis, C. C. (2020). Measuring readiness for implementation: A systematic review of measures' psychometric and pragmatic properties. *Implementation Research and Practice*, *1*, 1–29. https://doi.org/10.1177/2633489520933896

Zwerver, F., Schellart, A. J., Anema, J. R., & van der Beek, A. J. (2013). Changes in insurance physicians' attitudes, self-efficacy, intention, and knowledge and skills regarding the guidelines for depression, following an implementation strategy. *Journal of Occupational Rehabilitation*, *23*(1), 148–156. https://doi.org/10.1007/s10926-012-9378-9
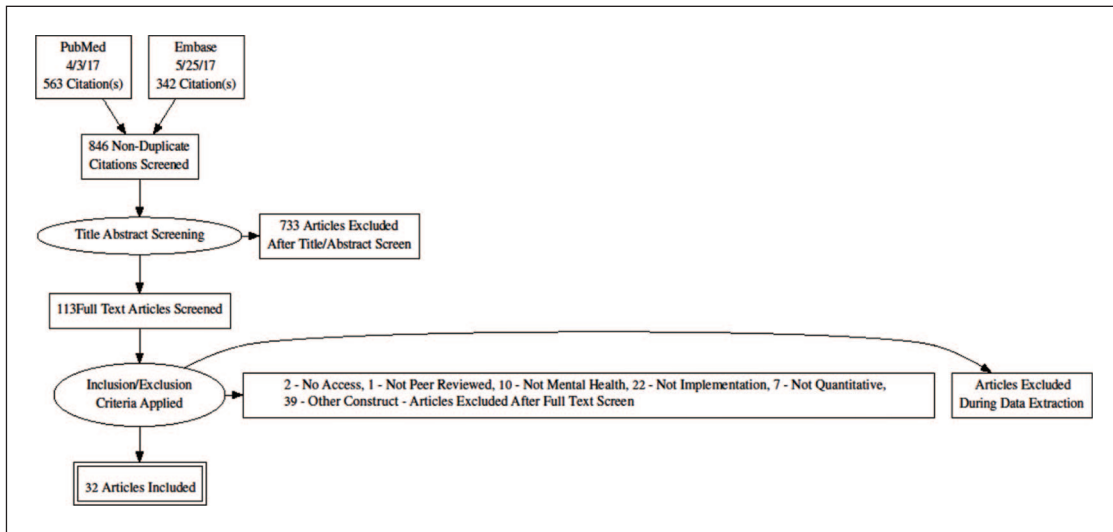
Appendix I

PRISMA diagrams

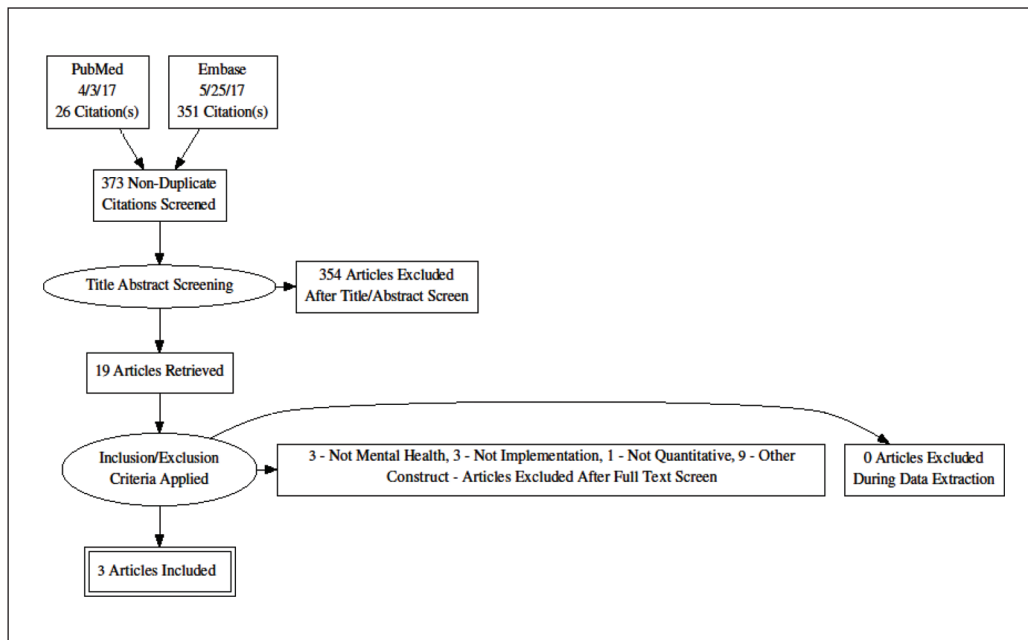**Figure 6.** Knowledge and beliefs about the intervention.
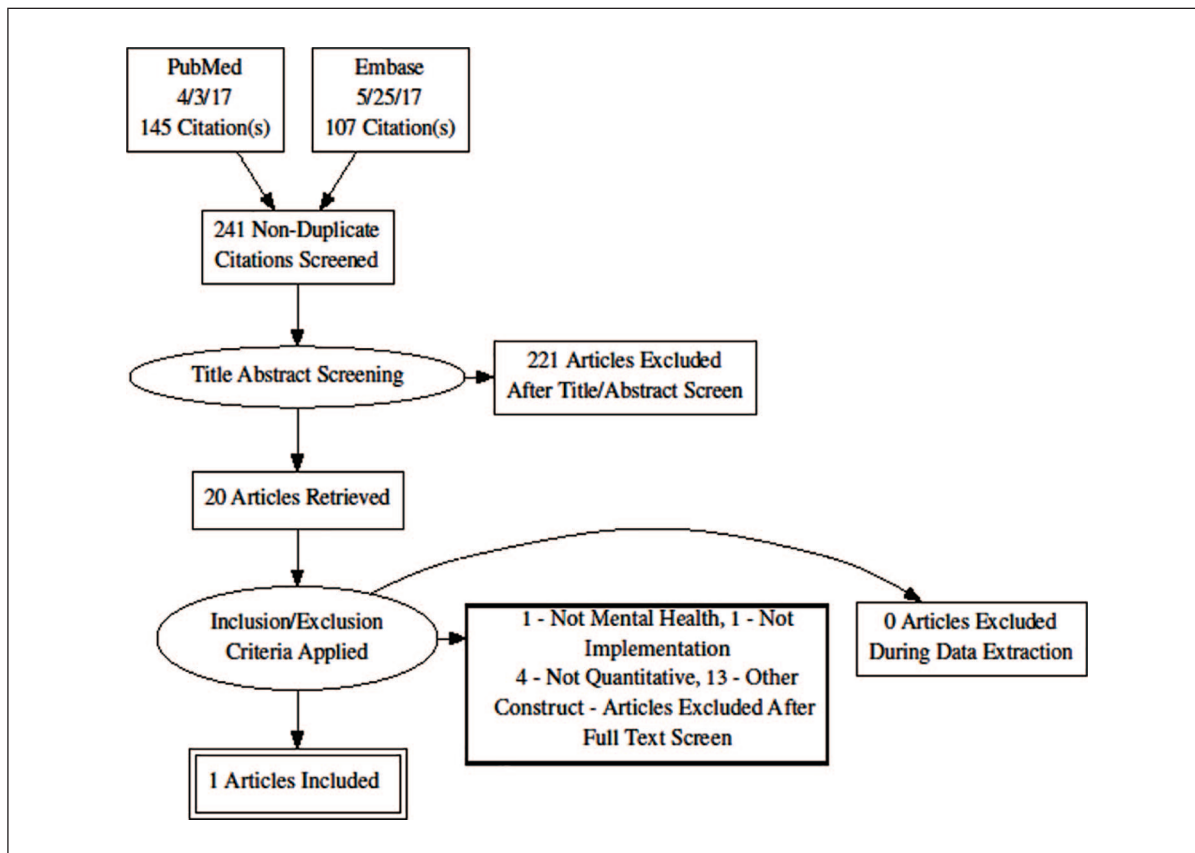

Figure 7. Self-efficacy.
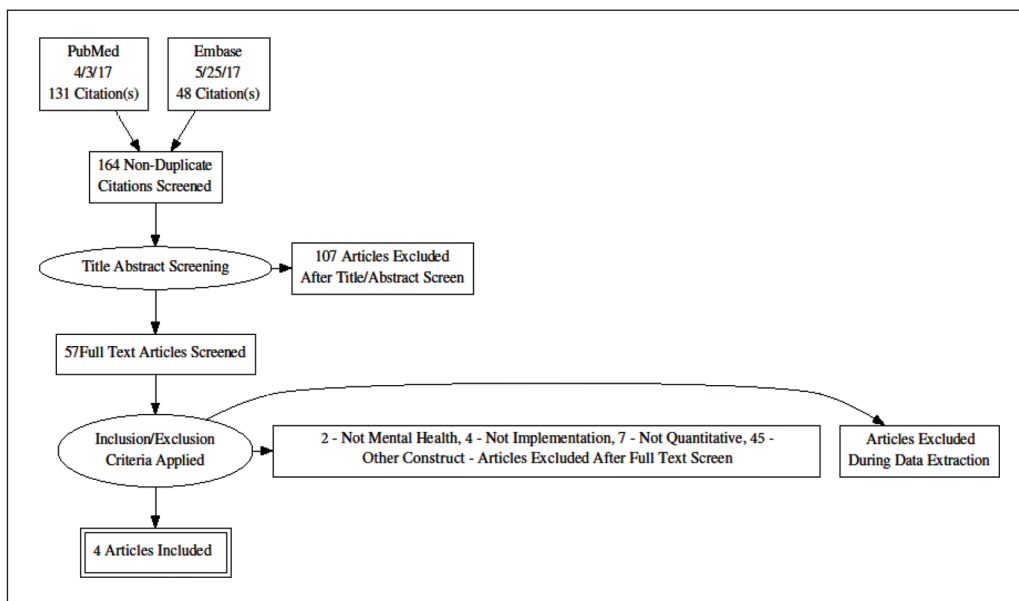
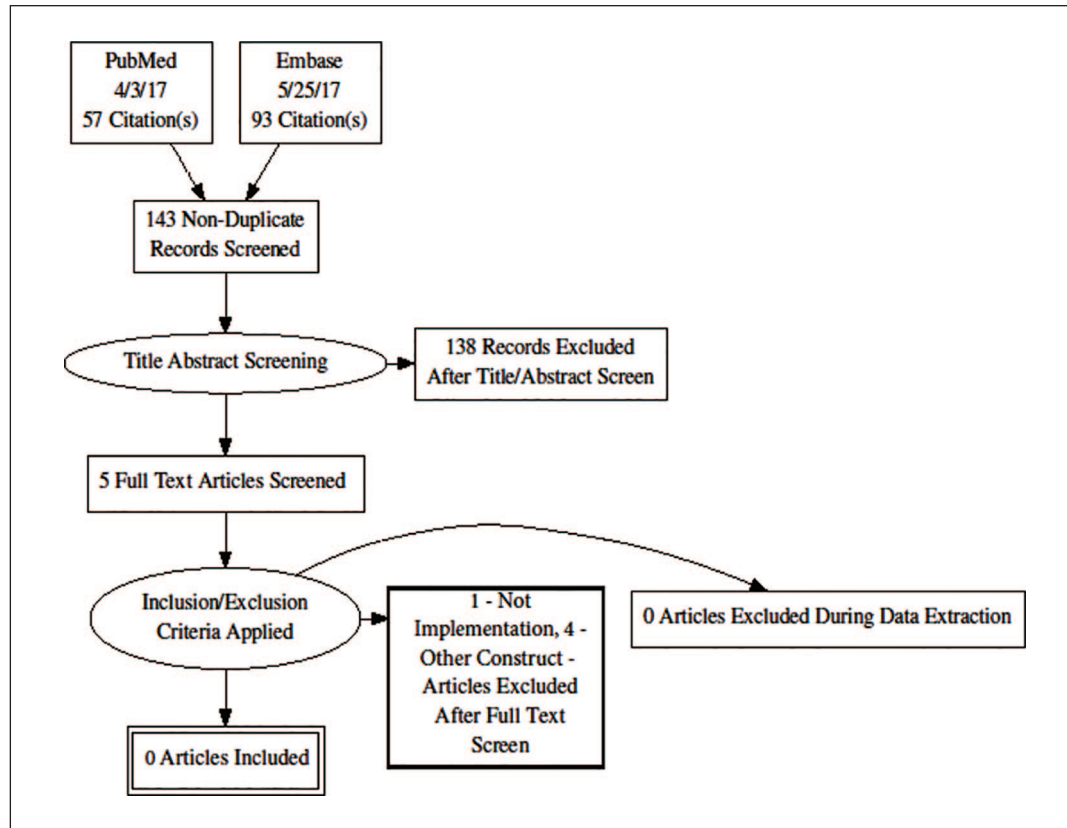**Figure 8.** Individual stage of change.

Figure 9. Individual identification with the organization.

Figure 10. Other personal attributes.